Strategies for Finding Disease Genes

Dennis Drayna

NIDCD

Human Genetics =

Study of Human Variation

Determining whether variation is genetic

- twin studies
- adoption studies
- segregation analysis

Simple Traits <- ---- -> Complex Traits

rare, Mendelian common diseases

penetrance

expressivity

multiple genes

gene-gene interaction

environment

Linkage studies

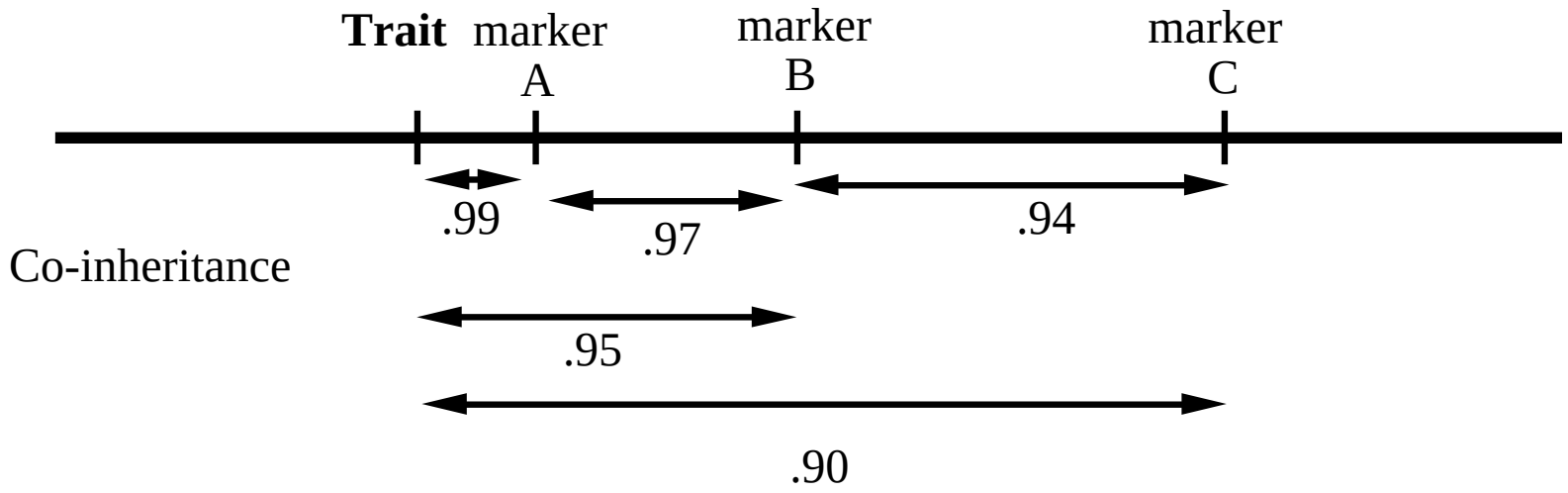
- Linkage- observance of co-inheritance of 2 traits when passed down from parent to offspring

  - violations of Mendel's 2nd law

Linkage

- ABO and Nail-Patella Syndrome
- percentage of co-inheritance corresponds to genetic (and physical) distance – a map!

A Genetic Map



Principle of positional cloning

- If you know the location of a gene which specifies a trait, you can identify that gene, without knowledge of the biochemistry, physiology, or pathology of the gene product
- Since we know the location of our genetic markers, observation of linkage tells the location of the gene

Genetic markers
DNA-based makers

- simple sequence repeats
  - CA's, tri-and tetra-nucleotide repeats
  - differences based on length across repeat
  - gttatcttagggctcagtcacacacacacacacacacacatccaggtattggatcaact
  - single copy variable repeat single copy
- SNP's - single nucleotide polymorphisms
  - ggattacctgaccctgAccgcttaatcattgatt
  - ggattacctgaccctgGccgcttaatcattgatt

Simple sequence repeats

- Use PCR, followed by gel electrophoresis, typically on automated (ABI) machine
- highly polymorphic
- frequently informative

Single nucleotide polymorphisms SNP's

- Use PCR, followed by other, non-gel analysis method
  - DNA chips, many other methods
- highly automatable
- Only 2 alleles at each locus
- Need more of them to make up for lower informativity

Doing a linkage study

- Assemble families
  - 10^1 - 10^3 individuals required
  - DNA - blood is traditional, other sources gaining popularity
  - Ongoing contact with families is essential
- Genome-wide search
  - 375 simple repeat markers for a first pass
  - A lot more SNP's

Doing a linkage study
Statistical Analysis

- LOD score method
  - used for Mendelian traits
  - Logarithm of the odds that: markers are linked, at a particular distance, divided by the odds that they're linked at 50% co-inheritance, i.e., they're not linked at all

LOD score table

θ

	0.00	0.01	0.05	0.10	0.20	0.30	0.40
family 1	- ∞	0.91	1.33	1.87	1.60	1.10	0.88
family 2	2.33	2.18	1.99	1.66	1.41	1.22	1.04
total	- ∞	3.09	3.29	3.53	3.01	2.32	1.92

LOD scores

- Historically defined
  - LOD of 1 = suggestive
  - LOD of 2 = probable
  - LOD of 3 = proof

Linkage Analysis

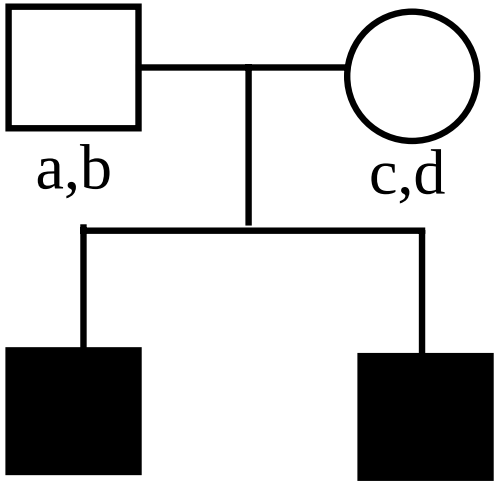
- Computerized statistical packages
  - LIPED
  - LINKAGE
  - MAPMAKER
  - FASTLINK

Non-parametric analyses

- Used for non-Mendelian traits
- Affected sib-pair, affected relative pair methods and allele sharing expectations
- Transmission disequilibrium test
- Support measured in p values, t values, other units
- GENEHUNTER

Allele sharing expectations

If Unlinked



No alleles	a,d b,c	c,b a,d	25% of the time
1 allele	a,d a,c a,d b,c	b,d b,c a,c b,d	50% of the time
2 alleles	a,d b,c	a,d b,c	25% of the time

What does the observation of linkage tell you?

- A gene contributing to the trait lies somewhere in the vicinity of the markers that show linkage
- Highest LOD score is never the marker that's closest to the gene
- Taken a 5 order of magnitude problem down to a 2 order of magnitude problem

Narrowing the location of the disease gene

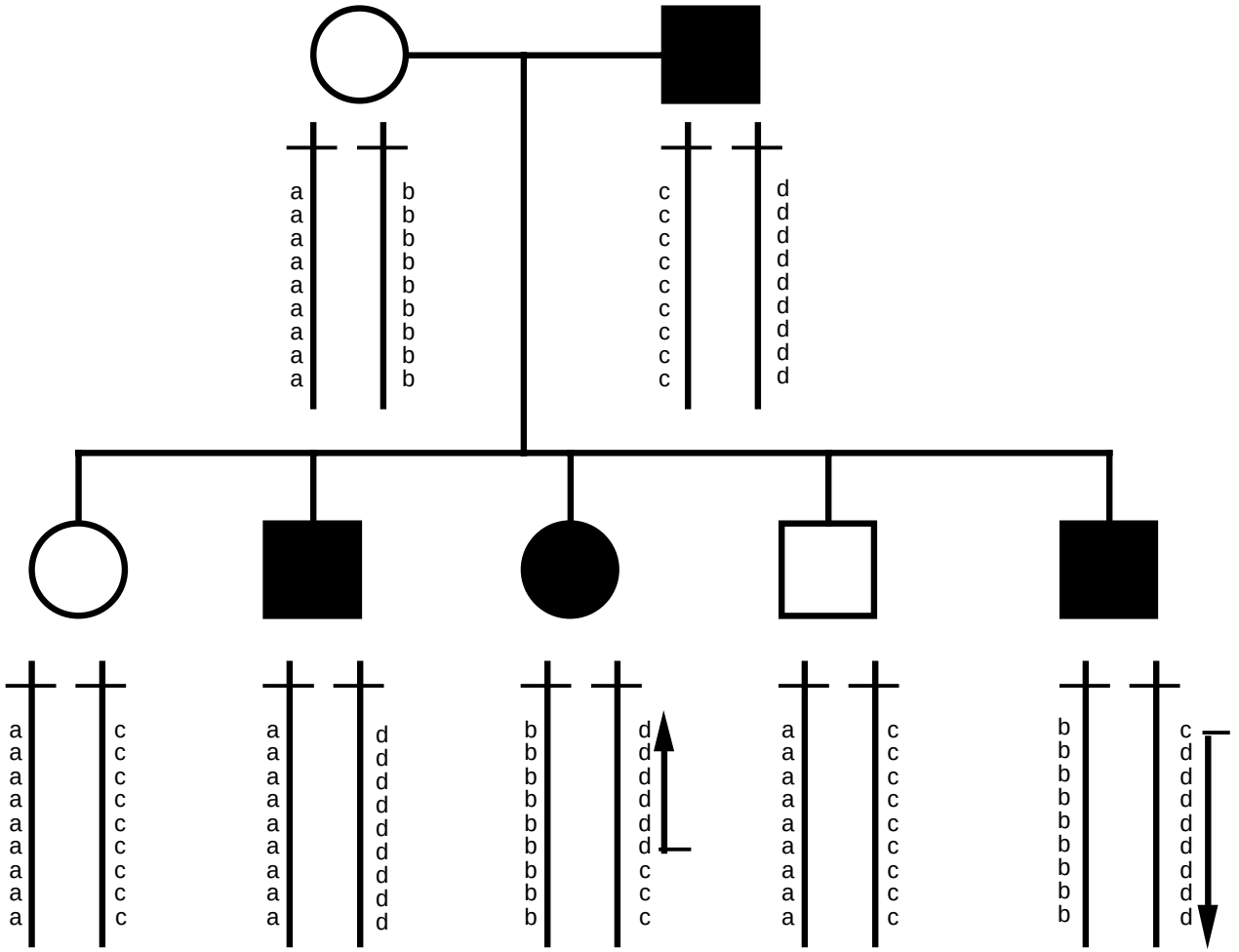
- Assembling a physical map of the region
  - BAC contig
- Radiation Hybrid (RH) mapping
- Construction of haplotypes and observation of cross-over events in families
  - Mendelian traits

Using haplotypes to find cross-overs

- Haplotypes - the arrangement of alleles on the 2 chromosomes within an individual

- Allow precise estimates of the region of the chromosome which segregates with the disease gene in families

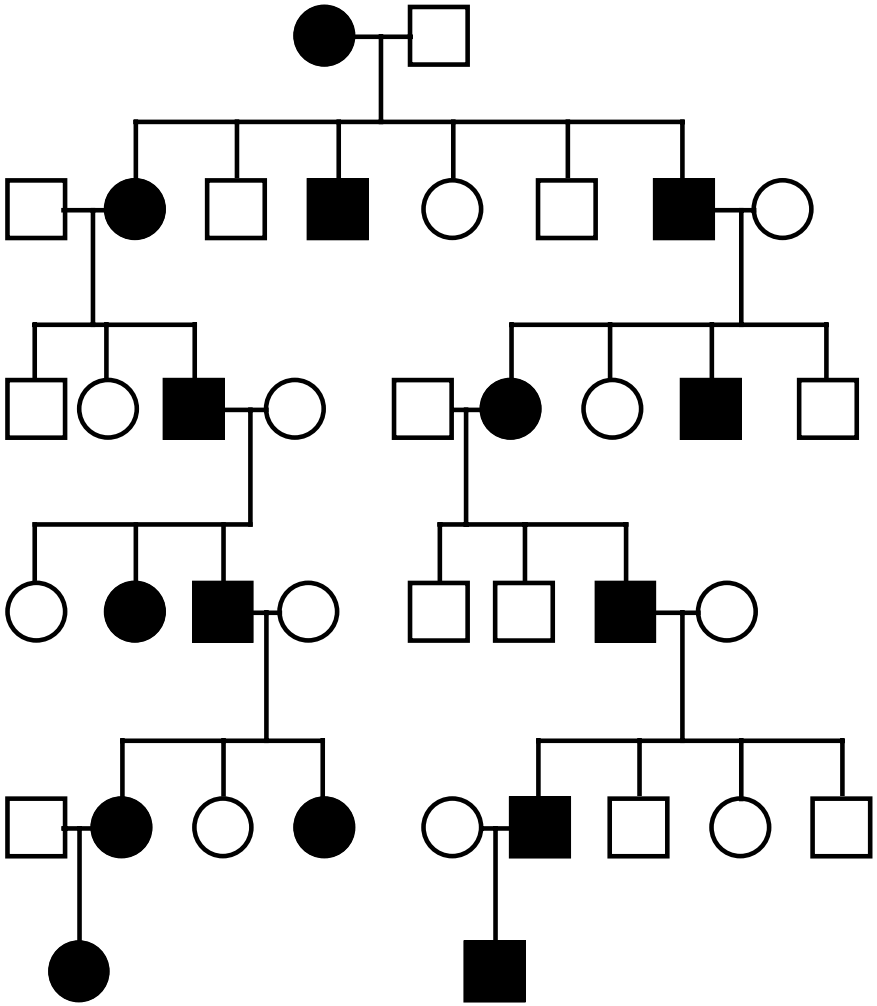
Using Haplotypes to find crossovers



Narrowing the location of the disease gene

- Using specialized populations to obtain more meioses
  - Founder effects, genetic isolates, inbreeding
  - Allele-sharing and linkage disequilibrium

Genetic Isolates



Candidate genes

- Biological plausibility has not been a good predictor
  - Failures frequently unreported
  - many disease genes turn out to be a surprise

Candidate genes

- Enumeration of all genes in the region
- now done electronically
  - NCBI

Candidate genes

- Evaluation involves comparison of DNA sequence between affected and unaffected individuals
- Genomic DNA sequence required
  - Mutations can exist in coding sequence, in introns, especially near splice sites, in promoter, in 3’ untranslated region

The Big Shortcut

- Large-scale rearrangements
- In rare patients, chromosome rearrangements can be observed in light microscope analysis -often associated with syndromic presentation
- Can exist at a scale larger than the entire gene, but too small to be visible in the light microscope

How do you know you've identified the real disease gene?

- Observation of different mutations in the same gene in different families
- Expression patterns
- Functional studies

Association studies

- Measure association of a marker with a disease in a population of unrelated affected individuals
- For markers not in genes, requires linkage disequilibrium – currently a risky strategy
- For markers that represent functional changes in genes (e.g. cSNP's), optimism is high

References - Textbooks

- T. Strachen and A.P. Read *Human Molecular Genetics 2*,Wiley-Liss, New York 1999 ISBN 0-471-33061-2

- J. Terwilliger and J. Ott *Handbook of Human Genetic Linkage* Johns Hopkins University Press, Baltimore 1994 ISBN 0-8018-4803-2

- J. L. Haines and M.A. Pericak-Vance, eds. *Approaches to Gene Mapping in Complex Human Disease*, Wiley-Liss, New York, 1998 ISBN 0-471-17195-6

References - URL's I.

- Online Mendelian Inheritance in Man - http://www.ncbi.nlm.nih.gov/Omim/
  - Human inherited diseases, medical genetics

- Genome Data Base - www.gdb.org
  - Genetic maps, polymorphic marker information, genetic disease information

- Marshfield Center for Medical Genetics - http://research.marshfieldclinic.org/genetics/
  - Genetic maps, marker screening sets, genotyping service

- Genethon - http://www.genethon.fr/genethon_en.html
  - Genetic maps, polymorphic maker information

References - URL's II.

- Center for Inherited Disease Research - http://www.cidr.jhmi.edu/
  - Genotyping service, data analysis
- Laboratory of Statistical Genetics at Rockefeller University - http://linkage.rockefeller.edu/
  - Linkage analysis programs, training courses, linkage bibliography and resources
- Single Nucleotide Polymorphisms -
  - http://snp.csl.org
  - http://www.ncbi.nlm.nih.gov/SNP/
  - Polymorphisms, locations, PCR primer sequences, allele frequencies, contacts to submitters
- The Golden Path - http://genome.cse.ucsc.edu/
  - Assembled human DNA sequence

References - URL's III.

- National Center for Biotechnology Information
- Genome Central - http://www.ncbi.nlm.nih.gov/genome/guide/central.html
- Genome sequence - http://www.ncbi.nlm.nih.gov/genome/seq/HsHome.shtml
- Gene Map 99 - http://ncbi.nlm.nih.gov/genemap99 - RH mapping
- dbEST - http://www.ncbi.nlm.nih.gov/dbEST/
- UniGene - http://ncbi.nlm.nih.gov/UniGene/
- Genome maps - http://genome.wustl.edu/gsc/human/Mapping/
- Ensembl - http://www.ensembl.org
- Oak Ridge Genome Channel - http://compbio.ornl.gov/tools/channel/